Iniziato	Thursday, 13 January 2022, 15:11
Stato	Completato
Terminato	Thursday, 13 January 2022, 15:38
Tempo impiegato	26 min. 16 secondi
Punteggio	15,00/15,00
Valutazione	**30,00** su un massimo di 30,00 (**100**%)

Domanda **1**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the main reason for the *standardization* of numeric attributes?

Scegli un'alternativa:

a. Remove non-standard values

b. Map all the nominal attributes to the same range, in order to prevent the values with higher frequency from having prevailing influence

c. Map all the numeric attributes to a new range such that the mean is zero and the variance is one.

d. Change the distribution of the numeric attributes, in order to obtain gaussian distributions

Domanda **2**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the two binary vectors below, which is their similarity according to the Jaccard Coefficient?

a b c d e f g h i j

1 0 0 0 1 0 1 1 0 1

1 0 1 1 1 0 1 0 1 0

Scegli un'alternativa:

a. 0.1

b. 0.5

c. 0.375

d. 0.2

Domanda **3**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the two binary vectors below, which is their similarity according to the Jaccard Coefficient?

a b c d e f g h i j

1 0 0 0 1 0 1 1 0 1

1 0 1 1 1 0 1 0 1 0

Scegli un'alternativa:

a. 0.2

b. 0.1

c. 0.375

d. 0.5

Domanda **4**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

In which mining activity the *Information Gain can be useful?*

Scegli un'alternativa:

a. Discretization

b. Clustering

c. Classification

d. Discovery of association rules

Domanda **5**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

When developing a classifier, which of the following is a symptom of overfitting?

Scegli un'alternativa:

a. The error rate in the test set is much smaller than the error rate in the training set

b. The precision is much greater than the recall

c. The error rate in the test set is much greater than the error rate in the training set

d. The error rate in the test set is more than 30%

Domanda **6**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

In a decision tree, the number of objects in a node...

Scegli un'alternativa:

a. ...is smaller than or equal to the number of objects in its ancestor

b. ...is bigger than the number of objects in its ancestor

c. ...is smaller than the number of objects in its ancestor

d. ...is not related to the number of objects in its ancestor

Domanda **7**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the main purpose of *smoothing* in Bayesian classification?

Scegli un'alternativa:

a. Dealing with missing values

b. Classifying an object containing attribute values which are missing from some classes in the test set

c. Classifying an object containing attribute values which are missing from some classes in the training set

d. Reduce the variability of the data

Domanda **8**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the statements below about *Hierarchical Agglomerative Clustering* is true?

a. Requires the definition of *distance between sets of objects*

b. Is based on a well founded statistical model

c. Requires the definition of *Inertia* of clusters

d. Is very efficient, also with large datasets

Domanda **9**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the statements below is true? (Only one)

Scegli un'alternativa:

a. Sometimes k-means stops to a configuration which does not give the minimum distortion for the chosen value of the number of clusters.

b. K-means finds the number of clusters which gives the minimum distortion

c. K-means always stops to a configuration which gives the minimum distortion for the chosen value of the number of clusters.

d. K-means works well also with datasets having a very large number of attributes

Domanda **10**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following characteristic of data can reduce the effectiveness of DBSCAN?

Scegli un'alternativa:

a. All the variables are the same range of values

b. Presence of clusters with different densities

c. Clusters have concavities

d. Presence of outliers

Domanda **11**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Match the rule evaluation formulas with their names

$$sup(A \cup C) - sup(A)sup(C)$$

$$\frac{1 - sup(C)}{1 - conf(A \Rightarrow C)}$$

$$\frac{conf(A \Rightarrow C)}{sup(C)}$$

$$\frac{sup(A \Rightarrow C)}{sup(A)}$$

Domanda **12**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Consider the transactional dataset below

ID	Items
1	A,B,C
2	A,B,D
3	B,D,E
4	C,D
5	A,C,D,E

Which is the *confidence* of the rule A,C ⇒ B?

Scegli un'alternativa:

a. 20%

b. 40%

c. 100%

d. 50%

Domanda **13**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following ***is not*** an objective of feature selection

Scegli un'alternativa:

a. Select the features with higher range, which have more influence on the computations

b. Avoid the *curse of dimensionality*

c. Reduce time and memory complexity of the learning algorithms

d. Reduce the effect of noise

Domanda **14**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What is the **coefficient of determination R^2**?

a. Measure the amount of error in a linear regression model

b. Provide an index of goodness for a linear regression model

c. Measure the amount of error in a regression model

d. An index of goodness for a classification model

Domanda **15**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the purpose of discretisation?

Scegli un'alternativa:

a. Reduce the number of distinct values in an attribute, in order to put in evidence possible patterns and regularities

b. Increase the number of distinct values in an attribute, in order to put in evidence possible patterns and regularities

c. Reduce the number of distinct values in an attribute, in order to increase the efficiency of the computations

d. Reduce the range of values of a numeric attribute, to make all the attributes more comparable

Vai a...

Machine Learning - Python Lab ►